

From air to insight: the evolution of airborne DNA sequencing technologies

Mia F. G. Berelson¹, Darren Heavens¹, Paul Nicholson², Matthew D. Clark³ and Richard M. Leggett^{1,4,*}

Abstract

Historically, the analysis of airborne biological organisms relied on microscopy and culture-based techniques. However, technological advances such as PCR and next-generation sequencing now provide researchers with the ability to gather vast amounts of data on airborne environmental DNA (eDNA). Studies typically involve capturing airborne biological material, followed by nucleic acid extraction, library preparation, sequencing and taxonomic identification to characterize the eDNA at a given location. These methods have diverse applications, including pathogen detection in agriculture and human health, air quality monitoring, bioterrorism detection and biodiversity monitoring. A variety of methods are used for airborne eDNA analysis, as no single pipeline meets all needs. This review outlines current methods for sampling, extraction, sequencing and bioinformatic analysis, highlighting how different approaches can influence the resulting data and their suitability for specific use cases. It also explores current applications of airborne eDNA sampling and identifies research gaps in the field.

INTRODUCTION

Airborne microbes exert several important influences over our lives, including the spread of human, animal and plant diseases [1–4] and even contributing to weather events as airborne microbes influence cloud formation by acting as ice-nucleating particles [5]. Airborne eDNA can come from multiple sources, including bacteria, fungi and viruses in the air, alongside additional taxa such as vertebrates (as they shed skin etc.) [6] and plants releasing pollen or spores [7].

Airborne eDNA exhibits temporal variation on both short and long timescales, ranging from hourly fluctuations [8] to seasonal changes [9, 10]. Given the dynamic nature of airflow and the influence of land use on airborne eDNA [11], it is likely that eDNA varies not only over hours and days but also over seconds, minutes, decades and centuries. Spatial variation also occurs in airborne eDNA, both vertically within the stratosphere [12, 13] and horizontally across cities and continents [10, 14], with species identified often indicative of land use [15].

Earlier airborne microbiota studies relied on passive spore traps, and pollen and fungi were identified through culturing and microscopy. However, these methods were labour-intensive and unable to capture the full diversity of organisms present. Recent advances have led to the development of devices capable of sampling air at rates of thousands of litres per minute (e.g. SASS 4100), greatly increasing the amount of airborne particles captured. With improved samplers, DNA extraction and amplification methods, low-input sequencing library preparation and the high sensitivity of next-generation sequencing (NGS), it is now possible to sequence amplified DNA from just 60 s of collection [16]. This increased sensitivity enables faster methods and the identification of more species [17, 18]. As a result, publications on airborne eDNA have increased, with examples including tracking the spread of SARS-CoV-2 in hospitals [19], monitoring endangered species [6, 20], preventing forest disease outbreaks [21] and real-time pollen monitoring [8, 22].

Received 04 December 2024; Accepted 01 May 2025; Published 28 May 2025

Author affiliations: ¹Earlham Institute, Norwich Research Park, Norwich NR4 7UZ, UK; ²John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK; ³Natural History Museum, London SW7 5BD, UK; ⁴Centre for Microbial Interactions, Norwich Research Park, Norwich NR4 7UG, UK.

*Correspondence: Richard M. Leggett, richard.leggett@earlham.ac.uk

Keywords: air; eDNA; metabarcoding; metagenomics; microbiome; next-generation sequencing (NGS).

Abbreviations: AMR, antimicrobial resistance; ASVs, amplicon sequence variants; EC, electrostatic collector; eDNA, environmental DNA; HYSPLIT, Hybrid Single-Particle Lagrangian Integrated Trajectory; ITS, Internal Transcribed Spacer; LCA, lowest common ancestor; MAGs, metagenome-assembled genomes; NCBI, National Center for Biotechnology Information; NGS, next-generation sequencing; OTUs, operational taxonomic units; qPCR, quantitative PCR; SNP, single nucleotide polymorphism; WGS, whole-genome shotgun sequencing.

001564 © 2025 The Authors

 This is an open-access article distributed under the terms of the Creative Commons Attribution License. This article was made open access via a Publish and Read agreement between the Microbiology Society and the corresponding author's institution.

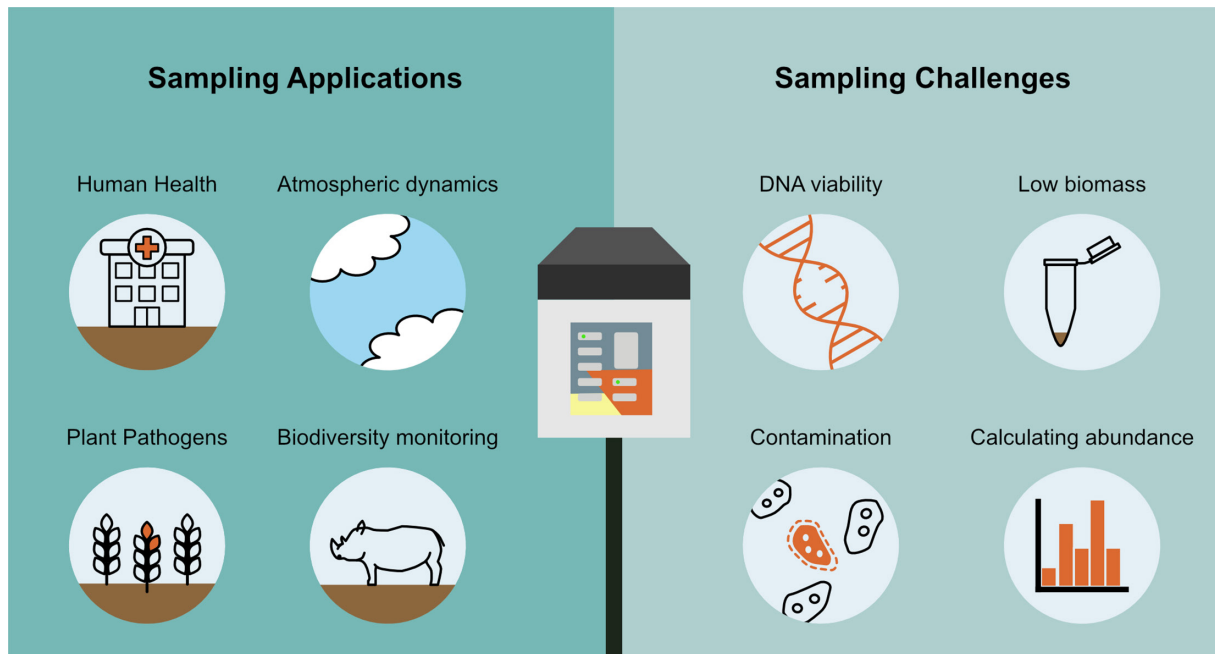


Fig. 1. Pictorial representation of applications and challenges of air sampling for eDNA analysis.

This review outlines the end-to-end process of airborne eDNA analysis, describing each stage from sampling to DNA extraction, sequencing and bioinformatic analysis. Furthermore, knowledge gaps, current applications and potential developments within the field are discussed (Fig. 1).

OVERVIEW OF AIRBORNE EDNA ANALYSIS

Airborne eDNA analysis typically involves four stages: sample collection, DNA extraction and isolation, amplification, sequencing and bioinformatic processing. The studies reviewed in this article follow this general pipeline (Fig. 2), though the specific method used at each stage can vary depending on factors such as study goals, available resources and equipment. For example, a study conducted in an area with low biomass may require longer sample collection (e.g. a week) to obtain sufficient DNA, whereas a study focused on tracking temporal changes to airborne eDNA in an urban setting might collect several short samples (ranging from 1 min to 1 h) over a single day.

Due to the complex nature of eDNA processing and analysis, the methods employed at each stage will introduce biases in the detection, identification and quantification of taxa. It is crucial to consider these biases when comparing airborne eDNA studies and to account for them before concluding.

Air collection

Air collection is the first critical step, as enough DNA must be obtained for sequencing and it needs to be sufficient to accurately represent the true composition of the air microbiome. Factors such as the sampler used, collection duration, flow rate and sampler height can all influence the detected airborne eDNA. These variables are interconnected, as samplers have limited collection times and flow rates, and their efficiency can vary.

Type of sampler

Air samplers primarily operate through impaction, impingement or filtering [23]. They can be either passive, with no air flow, or active, where air is drawn into the sampler. The advantages and disadvantages of these mechanisms are outlined in Table 1. Samplers differ in efficiency, flow rate and collection surfaces, all of which can influence the type and diversity of organisms detected. Consequently, the choice of a sampler should be tailored to the specific research question. Additionally, the sampler attributes can affect downstream DNA extraction and sequencing; active samplers with high flow rates will collect more DNA.

Since this review focusses on providing an overview of approaches that combine DNA sequencing with air sampling to identify airborne taxa, we do not extensively discuss the merits of different sampler models. Instead, we refer the reader to existing studies

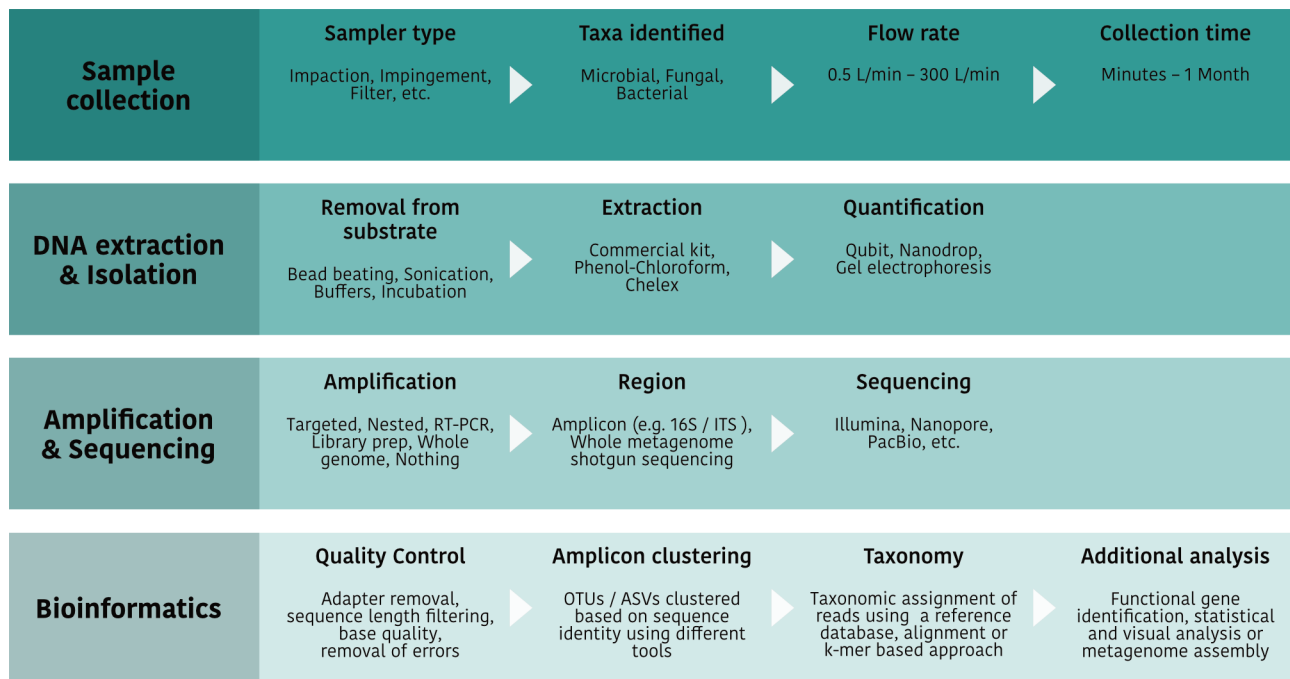


Fig. 2. Airborne eDNA analysis pipelines typically comprise four stages: sample collection, DNA extraction and isolation, amplification and sequencing and bioinformatics. Each stage includes several steps with different options depending on the study's research objectives.

that compare air sampler attributes [23, 24] and those focused on the specific sampling mechanism, such as passive samplers [25] and rotating arm impaction [26]. Below, we offer some general observations on the attributes of the three types of samplers.

Air sampler comparison

The ideal air sampler would capture 100% of the particles in the air that pass through its collector, representing the full diversity of airborne eDNA with enough sensitivity to detect even the rarest taxa. It would also allow tuning of the captured particle sizes according to the specific requirements of the experiment. However, such a device does not yet exist. Therefore, it is crucial to evaluate commercially available samplers based on their particle collection efficiency, measured eDNA diversity and sensitivity. Collection efficiency can be discerned by assessing the DNA concentration of a sample or, in the case of bacteria, by evaluating the 16S rRNA gene copy number. The ability of a sampler to fully capture organism diversity, including rarer taxa, is determined after DNA sequencing during the analysis phase.

One study comparing the collection efficiency of liquid impingement and filter impaction samplers used 16S rRNA gene copy abundance to assess DNA yield. The results showed that filter impaction recovered over an order of magnitude more gene copies

Table 1. Common commercially available air samplers, after [23, 24]

Category	Sampler available	Description	Advantage	Disadvantage
Impaction	Burkard 7 Day sampler, Andersen sampler, MicroPEM, AirPort MD8 364 device	Particles are impacted onto a surface (adhesive, agar, membrane filter etc.) transverse to airflow	Can select which particle size is recovered by adjusting the flow rate Direct collection onto agar for viability testing	Particle bounce can reduce collection efficiency
Impingement/cyclone	Bertin Coriolis, SKC Biosampler, VIVAS	Particles become suspended in a cyclone of air/liquid	Higher flow rates than filter samplers	Lower collection efficiency compared to filter samplers
Filtering	Thermo Scientific MFC-PM10 High Volume Air Sampler, Innovaprep Bobcat/Cub, SASS 3100/4100 Dry Air Sampler	Air is passed through a filter (fibrous, membrane or flat), which may include electrostatically charged filters ('electret')	Easy-to-use filter and recovery system Electret filters can collect particles smaller than the pore size	The choice of filter size determines collection efficiency for various particle sizes Not all particles captured (could also be an advantage)

than impingement [24]. This higher DNA yield is likely due to the increased particle retention efficiency of filter samplers. Another factor influencing sampler yield may be the effect of collection surfaces on downstream processes. In a separate experiment, researchers compared the DNA yield from culture and non-culture collections using membrane filtration, liquid impingement and an electrostatic collector (EC). The EC was the most efficient method, likely due to the much higher flow rate, i.e. 10–100 times greater than the other samplers tested [27].

Even when using the same sampler, the choice of capture material can affect collection efficiency. A comparison of five different membrane filters revealed significant differences in DNA recovery efficiency, with sequencing results showing variations in the diversity of recovered taxa [28]. Another study compared microbial communities detected by five air samplers and settled dust, finding that sampler type had a greater influence on microbial composition than collection location [29]. Differences in flow rates, particle size cut-offs, height and collection media likely contributed to this variation.

Studies on sampler sensitivity are less common in the literature, as they require controlled environments such as clean rooms. A comparison of two impingement samplers found that the VIVAS had higher sensitivity and detected viruses not observed in BioSampler collections [30]. Another study comparing eight samplers found that high flow rate samplers collected the most viruses, but lower flow rate samplers, including the VIVAS, more accurately determined airborne viral concentrations [31].

Research on bacterial pathogen detection also highlights differences between samplers. A study comparing impaction (AirPort MD8), impingement (BioSampler) and cyclone (Coriolis Micro) samplers found that all detected *Coxiella burnetii*, but their effectiveness varied at different concentrations [32]. Similarly, a study investigating airborne vertebrate DNA found that the Burkard spore trap recovered higher numbers of vertebrate species compared to cyclone samplers (Coriolis μ and Burkard multi-vial cyclone) [33].

Sampling volume

Collecting sufficient DNA is essential for downstream extraction, sequencing and analysis. However, collection duration and volume depend on experimental considerations, such as capturing all species present or monitoring temporal changes in airborne eDNA. Increasing air volume through a higher flow rate or longer collection time has been shown to increase DNA yield [24, 34] and likely increases the number of identified species, though published data on this relationship remain limited.

To comprehensively characterize all taxa present in airborne eDNA, collection periods may need to span days or weeks, particularly in low biomass locations. For example, in a study conducted at polar sites, samples collected for a fortnight contained 16S rRNA gene concentrations comparable to negative controls, suggesting that significantly longer collection times would be necessary for sufficient DNA recovery [35].

Species accumulation curves can help determine whether most taxa in a given area have been captured. The number of identified taxa generally increases with sampling time, reflecting the complex and dynamic nature of airborne eDNA. In one study, samples collected for 1–403 min showed no plateau in species richness when plotted as an accumulation curve [16]. As a result, studies aiming to capture the full diversity of airborne eDNA often employ 7-day collection periods [21, 36]. For studies investigating temporal changes in airborne eDNA, shorter ‘snapshot’ collections may be required. Higher flow rate samplers are particularly useful in these cases, as they can collect sufficient DNA within a short period. Notably, a personal impingement sampler has been developed that can detect microbes from just 10-s samples, albeit using quantitative PCR (qPCR) [37].

Sampler placement and environmental considerations

Choosing where to place the sampler is a critical decision in aerosol sampling experiments as factors such as height above ground, wind direction and surrounding habitat all influence what is detected. There is no universally optimal placement, but positioning can be guided by knowledge of the local area, meteorological parameters and the biological system of interest [38].

Indoor sampler placement depends on the study’s objectives. For instance, samplers may be placed in bedrooms to monitor overnight allergens, near infected patients to monitor airborne pathogens or in hospital corridors to study the spread of microbes of clinical concern [39–41]. A study comparing indoor samplers at different heights found no significant differences in microbiome composition or diversity, likely because indoor air is contained within the environment and relatively homogenous [40].

Outdoor samplers should also be positioned based on the study’s focus. For example, to detect spores from a soil-borne pathogen, samplers should be placed low to the ground [38]. Conversely, samplers positioned above the floral canopy are more likely to detect airborne pathogens arriving from distant locations [23]. One study comparing rooftop and in-field samples found that fungal species richness was higher in rooftop samples than in those placed above the crop canopy [42]. This is because samplers near the crop canopy primarily capture local spores, whereas those 10–30 m above ground collect a more mixed sample containing biological material from further afield [23].

Topography plays an integral role in airborne eDNA dispersion. Valleys and mountains influence wind speed and direction, affecting how biological particles travel. Large structures create ‘quiet zones’ where air movement is reduced; these zones can reach up to twenty times the height of the structure [38].

Time of sample collection

The timing of air collections is also extremely important, as both the time of day and season can significantly affect airborne eDNA composition. Knowing when a sample was collected is essential for accurate interpretation and comparison across airborne eDNA studies. Seasonal shifts in airborne eDNA are well-documented, with specific taxonomic groups more abundant at different times of year. For example, pollen from flowering trees and plants is more prevalent in spring, while grasses and mosses, along with wood-rotting fungi, are more abundant in summer. In autumn, fungal spores dominate, and winter is marked by reduced plant pollen and the presence of cold-tolerant bacteria [9, 43, 44].

In addition to seasonal changes, airborne eDNA composition is known to oscillate throughout the day with changes in UV, temperature and humidity. These environmental factors can create consistent daily patterns, such as with spore release. For example, powdery mildew is often abundant at midday, tropical fungi sporulate overnight (or before sunrise) and *Alternaria* spp. are more prevalent in the afternoon [45, 46]. Other shifts in composition may be more spontaneous. For example, rainfall can affect the composition of airborne bacteria and fungi by washing out existing particles while dispersing new microbial particles into the air through splash effects [47, 48]. High wind speeds increase microbiome diversity [49], while wind direction can affect the microbiota present [50] and humidity triggers spore release in many fungi [51].

Diurnal differences have also been observed, and one study within an urban environment found human pathogenic bacteria to be more abundant during the day than at night [52]. Another study in Siberia recorded that the fungal community has the highest relative abundance and diversity at night and decreases during the day, whereas the bacterial community peaked in the evening and decreased at night [43].

These dynamics highlight the importance of recording metadata – such as local weather conditions and the time of sample collection – to enhance data interpretation. Incorporating this metadata into analyses can reveal how different taxa respond to diverse environmental conditions, providing a more comprehensive view of airborne eDNA dynamics.

Particle size

Particle size is a key factor in airborne eDNA dispersal, influencing both residence time in the air and sampling efficiency. Airborne biological particles can be in the nanometre range, but typically range from 0.65 to 12 µm in many environments [53], with smaller particles remaining suspended longer and larger ones settling more quickly [54]. To improve capture rates, organisms with larger particles and shorter airborne times may require sampling closer to the source, for extended durations, or during periods when they are more likely to be airborne due to environmental or temporal factors. Additionally, different samplers are optimized for collecting specific particle sizes, making sampler selection critical. While literature on airborne particle size and eDNA monitoring is limited, Mahaffee *et al.* [38] provide a review on optimizing fungal spore detection, and several studies compare airborne community composition across different particle size fractions [48, 55–58].

DNA extraction and isolation

Following air collection, the next steps are to lyse cells and spores etc. from the samples, extract DNA and isolate the DNA from other biological matter – these processes can introduce bias into results. Most studies use a commercially available extraction kit with modifications such as additional incubation or filtering steps. Limited studies are comparing DNA extraction protocols for air samples, and they are often restricted to a specific sampler or collection medium.

Obtaining sufficient high-quality DNA from airborne samples is essential, as it affects sequencing, taxonomic assignment and other downstream analyses. Low DNA yields can result in poor sequencing data, while degraded DNA may lead to reductions in yield, misidentifications or increased errors. Because of this, it is important to choose an extraction method that maximizes DNA yield while preserving quality for accurate results.

Prior to DNA extraction, samples usually require concentrating either through filtration or centrifugation. A published comparison of filtration and centrifugation of collection liquid used the Internal Transcribed Spacer (ITS) gene to compare identified fungal communities. Filtration yielded three orders of magnitude more DNA than centrifugation and identified taxa absent in centrifuged samples [18]. These findings suggest filtration provides a more comprehensive fungal community than centrifugation for liquid samples.

Once concentrated, cells are lysed, and DNA is extracted, typically using commercially available microbiome extraction kits. However, none yet exist that are specifically designed for air samples (unlike, e.g. soil, for which specific kits exist). A comparison of the efficacy of four extraction kits targeted to specific sample types but tested on air samples collected onto quartz filters found that water and soil kits produced higher 16S rRNA gene copy numbers than blood or human tissue kits. Notably, the study did not conduct taxonomic comparisons to measure the sensitivity of the kits or the sample diversity captured [35]. An additional complication arises from contamination by the so-called kitome, which is the genetic material already present in these extraction kits, which may be misconstrued as DNA from the sample [59].

Cells of certain species are easier to lyse, more prone to degradation or less likely to adhere to tube walls, leading to their disproportionately high abundance in the DNA extract [18, 60]. For example, researchers found that DNA extracted with the Zymo Research Fungal/Bacteria DNA Microprep Kit from a mock community resulted in over- and under-represented species and that sporulating fungi were generally underrepresented [61]. Inevitably, bias is introduced during DNA extraction, and the observed taxonomic abundance is different from reality. Another study used captured aerosol samples to compare three extraction protocols with different incubation and lysis steps. The authors found that among the top 10 most prevalent species of each sample, only 5 were shared between the extraction methods, showing that the extraction protocol has a large influence on the species identified. This study also used centrifugation to concentrate the DNA and then separately analysed the pellet and supernatant, finding they had substantially different taxonomic composition [62].

These same studies compared the read numbers of different taxa from samples to highlight how bias can affect results, as different protocols may detect species at varying prevalence. While some level of bias is inevitable in metagenomic studies, minimizing it where possible is crucial. Equally important is maintaining consistency in sample handling and DNA extraction protocols throughout a project to ensure any remaining bias remains constant and does not introduce additional variability.

Other more comprehensive extraction protocols have been developed for metagenomic samples, such as the ‘three peaks’ faecal DNA extraction method, which uses three different lysis steps (chemical, enzymatic and mechanical) targeted towards different organisms within the sample while retaining long DNA molecules [63]. Although not routinely used on air samples, it has the potential to further optimize extraction by finding a better balance between lysis and DNA degradation.

Amplification and sequencing

After extraction, it is common for researchers to deploy metabarcoding, which involves PCR primer-based amplification of a barcoding gene or, less frequently, whole-genome shotgun sequencing (WGS), with or without whole-genome amplification. Whether to use metabarcoding or WGS depends on factors such as the required taxonomic resolution, time, cost and the specific research question. For example, a study investigating bacterial community composition may use 16S amplicon sequencing, while another looking at antimicrobial resistance (AMR) gene abundance would opt for WGS to capture a broad range of genes present in airborne eDNA.

PCR for known species

PCR approaches can be employed to target specific species, and this is particularly valuable in scenarios where only a few species are of interest and precise identification is crucial. For example, PCR has been used in the detection of *Campylobacter* spp. from air filters in poultry houses [2] and for the detection of multiple airborne microbes simultaneously from samples collected with a prototype personal bioaerosol sampler [37]. These methods are only able to detect specific species, but WGS or metabarcoding would be necessary to identify all potential taxa present (using reference genomes). Despite these limitations, PCR provides a rapid and specific method for detecting pathogens in air samples, thereby complementing broader metagenomic approaches, such as [2], where WGS was conducted alongside PCR.

Metabarcoding

Metabarcoding uses primers to target highly conserved genomic regions that contain short variable sequences, which can be used to distinguish taxa – sometimes to the level of species, but often to higher taxonomic levels or to operational taxonomic units (OTUs) representative of multiple species. For bacterial and fungal communities, this is often the 16S and ITS rDNA region, and amplicon approaches based on these regions have been successfully used to monitor airborne eDNA [11, 21, 24, 40]. Metabarcoding is often chosen as it is cheaper than WGS, and the use of PCR means it is more robust to low amounts of DNA [64]. However, PCR can impact the detected microbial diversity by inducing bias during amplification, while taxonomic identification from amplicons has lower resolution than with WGS data [34, 65]. Banchi *et al.* [61] have reviewed the use of metabarcoding in aerobiological studies in detail, so here we will discuss how amplification can affect the detected taxa. A comparison of the ITS1 and ITS2 regions for metabarcoding of airborne fungal spores found the fungal community composition to be different depending on the region amplified [66]. In addition, some species are only identifiable from one of the regions, for example, in fungi, *Armillaria* spp. only with the ITS1 marker and *Hymenoscyphus fraxineus* and *Melampsora larici-populina* only with the ITS2 marker [34]. Another study compared five different ITS primers with the same sample and detected different fungal communities with each set, with each primer pair able to detect ~50% of the fungal diversity which increased to 70–80%, when results from two different primer sets were combined [67]. In bacteria, a comparison of 16S and 12S amplicon sequencing identified different unique species [6]. Thus, airborne eDNA studies that use one metabarcoding region do not identify some unique species, limiting the characterization of airborne eDNA.

Further limitations of metabarcoding come from limited taxonomic resolution and bias in the detected community composition. Amplicon reads are often classified to higher taxonomic levels than species, and this has been demonstrated to be the case with fungi, bacteria and vertebrates [6, 40, 68].

As sequencing technologies advance, long-read metabarcoding is gaining traction, offering full-length marker gene sequences rather than shorter regions, thereby improving taxonomic resolution [69, 70]. The decreasing error rate of nanopore sequencing and the reducing cost of PacBio sequencing make these methods increasingly viable [71]. However, this data needs to be processed differently from other metabarcoding data [72]. There is currently a lack of published studies applying long-read metabarcoding to airborne eDNA, but it holds great potential for the future.

Bias introduced during amplicon sequencing

Biases arising from amplification within metagenomic studies are well documented in the literature [60, 73], with taxa being over- and under-represented in mixed samples after 16S PCR [60]. Several factors contribute to this bias, including primer choice, annealing temperature and the number of amplification cycles [73–77]. Increasing the number of PCR cycles has been shown to reduce community richness by ~4–10-fold, as taxa with more efficient primer binding sites are preferentially amplified, leading to an overrepresentation of some groups and the loss of others [75, 77].

Beyond amplification bias, copy number variation can also skew relative abundance estimates in amplicon sequencing because within a metagenomic sample, organisms have different numbers of the amplified gene. For example, in bacteria, the 16S gene copy number varies from 1 to 15 between species [78]. Organisms with a higher copy number will, therefore, have an inflated abundance in the sample.

Additionally, even single-base mismatches between primers and template DNA can skew amplification efficiency. These mismatches have been shown to cause preferential amplification of certain taxa, altering the relative abundances in the detected community. They can also affect taxonomic resolution by changing amplicon length, which impacts classification accuracy [74].

Although most metagenomic studies addressing these issues have focused on microbial mock communities or gut and stool samples, these findings are likely transferable to airborne eDNA studies. Given these challenges, it is unlikely that a single set of PCR conditions can amplify all organisms equally efficiently. For a more in-depth discussion of PCR biases in metabarcoding, readers are referred to recent reviews on this topic [79, 80].

Whole-genome sequencing

WGS can detect all kingdoms simultaneously and be used for species and functional gene identification. In the absence of other constraints, it may be considered a preferable method for profiling airborne eDNA as it identifies a broader range of species than metabarcoding [4, 39, 43, 62, 81–83].

Another advantage of WGS is that it typically requires little or no amplification, reducing biases associated with PCR. As a result, relative abundance estimates from WGS data are typically closer to the true community composition than with amplicon sequencing. However, a drawback of using WGS over amplicon approaches is that it requires a larger quantity of starting DNA, which can be a limitation when working with air samples, which typically have low biomass. One approach to mitigate this challenge is to reduce reaction volumes during library preparation, as demonstrated in a study where a nanopore sequencing run was successfully conducted using just 6.25% of the recommended DNA input [84].

A study comparing 16S metabarcoding and WGS of air samples found both methods to converge at the phylum level but did not produce significant agreement at the species level, likely due to the 16S providing a less accurate result at the species level [34]. Additionally, WGS enables the simultaneous comparison of bacterial and fungal species, as both groups can be sequenced together, ensuring that their relative abundances are directly comparable. In contrast, using different primers for amplification in metabarcoding results in disparities that make abundance comparisons between fungi and bacteria more unreliable. A further benefit of WGS is the ability to assess the functional diversity of airborne eDNA – for example, to see the presence of detoxification, antibiotic resistance and virulence genes. However, a weakness of WGS is the less comprehensive reference databases that are currently available for classifying taxa – for example, the latest UNITE release (v10) contains ITS sequences representing ~11 times more fungal species than the number of fungal species with reference genomes available in the National Center for Biotechnology Information (NCBI) RefSeq (release 228).

Bioinformatics

In common with other metagenomic studies, there are typically two key questions that are asked of the data in an air eDNA study: ‘What is there?’ (taxonomic classification) and ‘What is it doing?’ (functional analysis). Here, we briefly outline the bioinformatic steps involved and how they have been applied to air samples. Fig. 3 summarizes these steps along with commonly used tools for each stage, references for which can be found in Table 2. For more detailed discussions on eDNA data processing and its impact on taxonomic and functional classification in amplicon and WGS studies, see [85–89].

Quality filtering

Quality filtering is a necessary pre-processing step and will help to prevent misidentification from alignment of low-quality reads with greater numbers of errors. Initially, sequencing adaptors are removed, and paired-end reads are merged where applicable

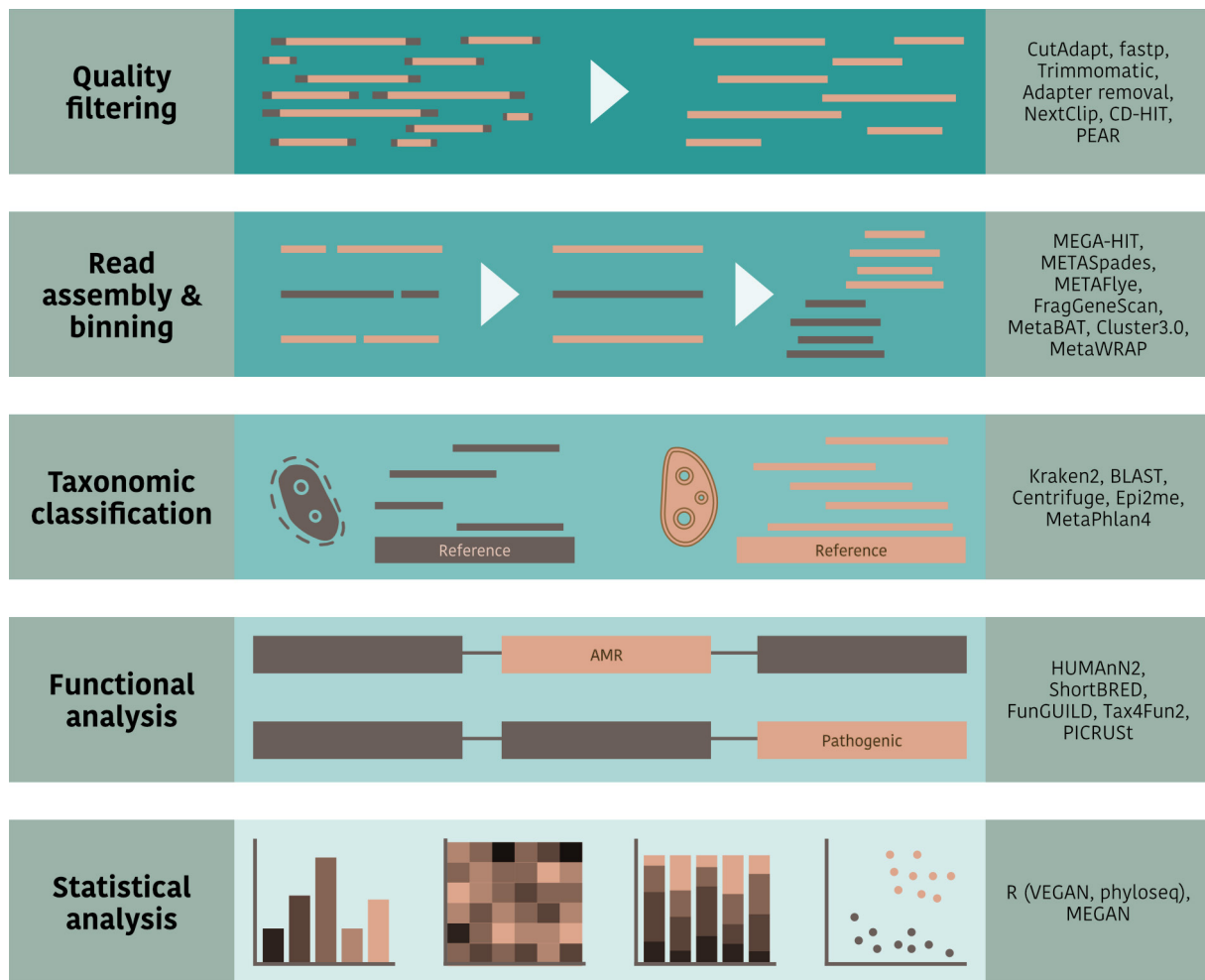


Fig. 3. Stages in bioinformatic analysis for sequence data generated from air samples. On the right of each row is listed a selection of commonly used tools. Further details of the tools can be found in Table 2.

and libraries with multiple samples are demultiplexed. Reads are then filtered on a variety of metrics but most often on length and quality score, the specific thresholds depending on the sequencing platform and study objectives. A review of the importance of quality filtering sequence data for downstream analysis is available at [90].

WGS read assembly

An additional optional step before taxonomic classification of WGS reads is assembly, in which shorter reads are combined into longer contigs to improve taxonomic precision by providing more genomic context, particularly for closely related species (see Ayling *et al.* [91] for a review). This approach is widely used in metagenomics and has been applied to airborne eDNA studies, enabling gene and protein prediction, and alignment to reference databases for fungal spore identification and species confirmation [4, 83, 92]. Following assembly, contigs may be grouped or ‘binned’ into genome-level metagenome-assembled genomes (MAGs) for further analysis. A large number of tools are available to facilitate this process, and in-depth discussions can be found in [93–95].

Amplicon binning

Metabarcoded samples are initially clustered into OTUs or amplicon sequence variants (ASVs) prior to taxonomic identification. OTUs are defined based on a sequence similarity threshold, which helps reduce the impact of sequencing errors but may also group closely related species together. In contrast, ASVs distinguish true biological variation from sequencing errors by incorporating sequence abundance, allowing for higher resolution at the cost of potentially discarding low-abundance sequences as noise.

While ASVs generally offer greater resolution, especially in terms of species-level identification, they can also be more susceptible to noise when dealing with rare taxa, which is particularly relevant in airborne eDNA, where low-abundance organisms may be

Table 2. Commonly used tools for bioinformatic processing. A more in-depth discussion can be found in [153] and in some of the review papers cited in the first column

Stage	Specific purpose	Tool name	Reference	
Quality filtering See also [90]	Remove adaptor sequences	CutAdapt	[154]	
		fastp	[155]	
		Trimmomatic	[156]	
		Adapter removal	[157]	
	Remove duplicated reads	NextClip	[158]	
		CD-HIT	[159]	
	Map reads to human reference and remove from sequences	Minimap2	[160]	
		BowTie	[161]	
	Merge paired-end reads	PEAR	[162]	
		Generate longer contigs from short reads to generate MAGs	MEGAHIT	[164]
metaSPAdes	[165]			
metaFlye	[166]			
metaMDGB	[167]			
Binning See also the discussion of binning tools for WGS [94] and amplicon [96]	Identify fragmented genes in short reads	FragGeneScan	[168]	
	Group contigs together based on sequence similarity	MetaBAT	[169]	
		MaxBin	[170]	
		CONCOCT	[171]	
		Cluster3.0/Pycluster	[172]	
		MetaWRAP	[173]	
Taxonomic classification See also review of classification tools [174]	Identify taxa from sequences by alignment to reference genomes	minimap2	[160]	
		BLAST	[97]	
		BWA	[175]	
	Identify taxa from sequences with a <i>k</i> -mer-based approach	Kraken2	[98]	
		Centrifuge	[176]	
		MASH	[177]	
	Identify taxa from sequences with a marker-gene approach	MetaPhlan4	[99]	
	Functional analysis	Identify genes with known functions	HUMAnN2	[178]
			ShortBRED	[179]
			FunGuild	[115]
Tax4Fun2			[180]	
PICRUSt			[181]	
Statistical analysis See also [121–123]	Perform abundance and diversity measures	VEGAN in R	[182]	
		Phyloseq in R	[183]	
Visualization	Plot graphs of the analysis	Ggplot2 in R	[184]	

Continued

Table 2. Continued

Stage	Specific purpose	Tool name	Reference
Multiple stages	Used in combination with BLAST to classify taxa and can be used to carry out statistical analysis and visualization	MEGAN and MALT	[185, 186]
	Bioinformatic workflows specific to Oxford Nanopore reads (basecalling, alignment and assembly)	Epi2me	https://labs.epi2me.io/
	Real-time classification and AMR analysis for nanopore sequencing	MARTi	[187]
	Amplicon-specific software tool that requires fastq files as inputs, which are then quality filtered, clustered into ASVs and assigned to taxa	DADA2	[188]
	A microbiome platform that is open-source, free and community-developed; the tool records the steps in bioinformatic pipelines to ensure they are reproducible and reusable by others	Quantitative Insights Into Microbial Ecology (QIIME 2)	https://qiime2.org/
	A tool designed to bin >100bp reads; can also assign taxonomy, build phylogenetic trees, perform functional analysis and compare datasets	MetaBin web server	[189]

underrepresented. For example, in airborne bacterial pathogen studies, assembly methods have successfully enhanced 16S read clustering into OTUs, improving species-level classification [52].

The choice of clustering approach (OTU/ASV) and parameters used will affect the community composition downstream analysis and is, therefore, an important consideration in study design. For example, a comparison of OTUs and ASVs clustering on the same dataset found little agreement at the species level but more consistency at the genus level [21], highlighting how the clustering approach can affect the result.

For a more detailed discussion of denoising and clustering methods, Hakimzadeh *et al.* [96] provide an in-depth analysis of these methodologies and their application to metabarcoding data.

Taxonomic classification

Clustered OTUs/ASVs or WGS reads are then mapped to reference databases for taxonomic identification, and the methods used for this vary. Historically, taxonomic identification was carried out with alignment-based approaches using BLAST [97] or other BLAST-like tools. However, as reference databases have expanded, alignment methods have become increasingly compute and time hungry, leading to many adopting more computationally efficient tools using *k*-mer-based strategies such as Kraken2 [98], or marker gene approaches like MetaPhlan4 [99]. These *k*-mer-based approaches typically require an up-to-date taxonomy of the reference database to be provided, and assignments may be inaccurate if this taxonomy is incorrect.

The choice of database is crucial for any alignment or *k*-mer-based analysis, as these approaches can only detect species that are represented in the database. If an organism is missing, reads may be misassigned to a closely related species, leading to false positives. Studies have shown that the detected community composition can vary greatly depending on the database used [100]. Many databases lack sequences for a substantial portion of microbiome diversity, particularly in airborne eDNA studies, where many microbial taxa are unculturable. This leads to species being classified at higher taxonomic levels or misassigned to closely related taxa. Errors in reference data, such as incorrect annotations, poor quality references or contaminated genomes, can further skew results.

For classification approaches utilizing lowest common ancestor (LCA) assignment (in which reads are placed at the lowest point in the taxonomy consistent with all 'good' matches), the increasing size of databases has been shown to increase the number of classified reads but make species-level taxonomic assignment more difficult. This occurs because the number of new species being uploaded to databases far outpaces the addition of new genera, increasing the likelihood of ambiguity at finer taxonomic levels. As a result, the LCA approach is more likely than it was to assign reads at the genus level rather than the species level [101]. Alongside this, the increasing size of databases makes greater demands on computing resources and makes it difficult to keep them up to date. For a more in-depth discussion of these reference databases' problems, see [102].

Airborne eDNA studies often map WGS reads to large databases like NCBI's nucleotide or non-redundant databases [6, 33, 39, 43, 44, 103], but some opt for smaller, curated databases tailored to specific taxonomic groups and research questions. Metabarcoding studies use gene-specific databases (e.g. ITS, 16S and 18S) for quicker analysis and fewer misalignments. For instance, the UNITE database [104], which only contains ITS sequences, has been used to study airborne *Alternaria* spp. [105], rice

pathogens [50] and fungal communities in forests [12]. In bacterial metabarcoding, the SILVA rRNA database, which includes 16S and 18S sequences [106], has been used to identify urban airborne bacteria in Tokyo [49], analyse live airborne micro-organisms [107] and identify airborne bacteria over the ocean [108].

Some study authors create custom databases, such as using viral reference genomes from NCBI to investigate seasonal changes in the airborne virome [109]. Similarly, the PHI-base (plant–host interaction) database [110] was used to generate a custom pathogen database in order to identify airborne plant pathogens [4].

Taxonomic assignments, particularly to species or strain level, should always be treated with a degree of caution. Post-filtering steps can be implemented to increase certainty. For example, a study of airborne eDNA in West Siberia reduced incorrect mappings by only accepting species-level classification when enough reads aligned uniquely to a single species [43]. This threshold will differ between studies, depending on factors like amplification and sequencing depth. Studies that do not amplify DNA may identify rare species with fewer reads than the above study, and disregarding these reads could result in the loss of measured diversity.

Functional analysis

Alongside taxonomic classification, functional analysis of airborne eDNA can provide insights into the activity of the organisms present. Metabarcoding genes can be used only to *infer* the function of present taxa, while WGS can be used to identify specific genes or transcripts. This means that WGS offers greater functional insights, assuming sequencing depth and DNA yield are high enough – though these both can be limiting factors in airborne eDNA studies

Functional analysis in air studies has largely focussed on the presence of AMR genes, using databases such as Comprehensive Antibiotic Resistance Database [111] or Structured Antibiotic Resistance Gene [112]. For example, studies have been conducted on the presence of airborne AMR genes using WGS on public transport [81] and hospitals [113] and real-time qPCR on samples collected from poultry sheds [114].

Functional analysis has also identified plant pathogenic potential in airborne fungi [21, 66] using the ITS region and tools like FunGuild [115]. Similarly, studies on airborne bacteria along the Antarctic coast used 16S rRNA gene analysis to reveal functional profiles related to biosynthesis and metabolism [116]. Additionally, a meta-analysis of global airborne microbiome diversity showed that anthropogenic activities correlate with increased abundance of pathogenicity genes in the air, using 16S data [117]. This aligns with a similar study in China, which found the relative abundance of pathogenic bacteria to increase as air quality decreased, using qPCR of specific genes [118].

There is also the potential to identify other functional genes such as virulence or toxicity within a metagenomic sample, as seen in marine and soil samples [119, 120]. But to our knowledge, there exist no published studies applying these approaches to air samples. Examining the functions of airborne microbes can reveal important insights into antimicrobial resistance, pathogenicity and metabolic potential, increasing knowledge on how these communities interact with their environments.

Statistical analysis

Once classified, data can be imported into analysis software like R or processed through custom bioinformatic pipelines to assess diversity, uncover patterns or dependencies or discover ecological insights relevant to environmental health and microbial dynamics. Read counts per taxa can be used to calculate alpha and beta diversity for a sample set. There are several reviews discussing the statistical analysis undertaken for microbiome data, which can also be applied to airborne eDNA studies [121–123].

Airborne eDNA studies typically involve other bespoke analyses, for example, detecting specific pathogens [50, 124, 125], correlation analysis with weather and environmental data [4, 49, 92] or determining how airborne eDNA composition changes with distance from a source such as a waste treatment plant [11] or from land [108].

CURRENT APPLICATIONS OF AIRBORNE SAMPLING

Sequencing of airborne eDNA has been applied to a vast range of fields encompassing plant and pathogen detection, high-altitude and oceanic air sampling, human health monitoring and biodiversity estimation (Fig. 1).

Detection of plants and their pathogens

Airborne eDNA analysis can be used to monitor plant pathogen outbreaks, including within agricultural or forestry systems. Historically, these settings have relied on passive collection setups in which airborne spores land on a sticky tape for analysis with microscopy. This is a slow and labourious process which requires identification expertise, but NGS promises a rapid, affordable and accurate airborne detection methodology.

A comparison of microscopy and metagenomics on the same spore trap samples showed that both methods were able to detect major cereal pathogen genera. However, the metagenomic approach detected pathogenic species that microscopy missed [92]. This demonstrates that DNA sequencing can enhance the knowledge gained from air sampling.

Molecular airborne eDNA techniques also enable large-scale monitoring without the need for labour-intensive surveys. For example, qPCR of known forest pathogens across 12 sites allowed for the detection of wind-dispersed diseases without the extensive effort of typical field surveys [21]. Similarly, a UK study used air sampling and WGS sequencing to identify several agriculturally significant crop pathogens [4]. In another example, airborne rice pathogens, disease severity and environmental conditions were monitored throughout a growing season using both qPCR and ITS metabarcoding [50].

The vast amount of data generated from airborne eDNA studies can be analysed in combination with climate and meteorological data to better understand airborne particulate movement. This approach enhances models of pathogen outbreaks [126] and helps explain how airborne biological particles interact with the atmosphere, leading to ice nucleation [127]. In some cases, the relationship between airborne eDNA and meteorological data can be analysed in reverse. For example, a global network of samplers revealed that annual mean temperature significantly affects spore community composition [10]. Samplers can also collect pollen from the air, as demonstrated in a study involving collection along a distance gradient from a GM maize field in which real-time qPCR was used to determine the ratio of GM to WT pollen and to monitor the dispersal of GM pollen [128].

High altitude and oceanic air sampling

Sampling airborne eDNA at high altitudes and over oceans presents new opportunities to investigate airborne microbial diversity and expand our knowledge of these environments beyond conventional studies.

The height above ground at which sampling occurs is known to affect the composition of airborne eDNA [129]. Most sampling is conducted near ground level, but some researchers have attempted to classify aerosol composition at much higher altitudes, up to 12,200 m above sea level. These collections require aircraft with air sampling devices that have been adapted for motion at high altitudes and are programmed to collect for specific times at different heights [58, 129–131].

Additionally, sampling at high altitudes indicates airborne eDNA composition differs with height above ground; one study found that at heights >1,000 m micro-organisms no longer had a diurnal cycle [129]. In-flight sampling has also been conducted above high-intensity forest fires and has shown that fires can increase microbial concentration and diversity of viable cells [132]. Variation in microbial composition with height needs to be considered when comparing different samples and experiments.

In addition to surveys in the sky, there have been studies of airborne eDNA over the ocean [15, 108]. For example, a study of the airborne bacterial community along the Antarctic coast found bacteria which contain genes likely to contribute to ocean-surface growth and survival [116].

Collectively, these findings underscore the necessity of broadening the scope of airborne eDNA research to encompass untapped environments that may exhibit different dynamics than those discussed. The insights gained from such investigations could enhance our understanding of the role micro-organisms play in ecological processes and atmospheric dynamics.

Sampling for human health

Air sampling has been used in hospitals for various purposes, including establishing microbiome community baselines. These baselines can help identify unusual increases in pathogen abundance in the future [39]. For example, with the emergence of SARS-CoV-2, air samples have been collected from around different parts of hospitals in order to pinpoint high-risk infection zones [1, 19, 41].

Wearable air samplers have been used to detect monkeypox in hospitals [133]. In this instance, healthcare workers were wearing PPE while in infectious rooms, but sampling devices could be used in the future to monitor workers' exposure to infectious diseases. Additionally, air samples taken from rooms of norovirus patients have helped researchers better understand how the virus is transmitted [55]. The same study also found that viable virus particles can survive in the air and linked their presence to how recently a patient had vomited.

While these examples mostly involved amplicon sequencing, whole-genome sequencing of air samples in hospitals and nearby areas has been used to identify the widespread presence of airborne antimicrobial resistance genes and to show that they are more abundant and diverse within hospitals than in nearby locations [113].

These applications illustrate the potential value of air sampling in healthcare settings, by monitoring pathogens, worker exposure and antimicrobial resistance.

Air sampling for biodiversity estimation

In recent years, molecular methods have become an important tool in the assessment of biodiversity. Airborne eDNA analysis offers a multitude of potential benefits: it will require less labour to monitor large areas, air sampling could occur in inaccessible areas using drones and species that are unlikely to be seen with conventional sampling may still be detected through their DNA. Several proof-of-concept studies have been published.

Vertebrate detection through air sampling has been demonstrated in a zoo with non-native species successfully detected up to hundreds of metres away and with such high sensitivity that the organisms present even in animal feed could be detected [6, 20]. Another study compared air samples collected with Burkard spore traps to images from camera traps and found that all species observed with the camera traps were identified using airborne DNA, as well as unique species only seen through air samples [33]. Air sampling in a botanical garden was able to identify 67 plant species out of the 1,585 known to be growing in the garden. The lack of identification of the other species could be due to the plants not releasing pollen at this time, but the authors also suggest that an increase in sampling effort would lead to a higher number of species being detected [7].

Beyond samples collected specifically for biodiversity assessment, filters intended for air quality monitoring have been successfully utilized to extract DNA. In one notable example, researchers measured biodiversity over a 34-year period (1974–2008) using weekly filters from a radionuclide monitoring station that each processed over 100,000 l of air. Deep shotgun sequencing detected over 2,700 genera and correlated land use changes with declines in biodiversity [44]. A similar study also used eDNA extracted from filters in air quality monitoring stations and identified over 180 taxa including vertebrates using amplicon sequencing [36]. These findings indicate that it is possible to assess historical biodiversity when filters are stored under appropriate conditions leveraging pre-existing air monitoring networks.

Collectively, these studies highlight the versatility of air sampling as a tool for biodiversity monitoring, enabling both real-time species detection and retrospective biodiversity assessments.

CURRENT RESEARCH GAPS

Is captured DNA viable?

Detecting an organism's DNA in the air does not necessarily indicate that the organism is still viable, which is an important factor in many airborne pathogen studies. In agricultural settings, costly fungicide applications may only be necessary if the detected spores are viable. Similarly, when studying human pathogens, public health recommendations could depend on whether airborne pathogens pose a real risk to people's health.

The current best method to assess spore viability is to collect directly onto a growth medium and count the colonies after incubation; the medium can be selective for a specific pathogen [134] or combined with sequencing to identify maximum taxonomic diversity and viability in tandem [58]. However, since culturable fungi make up less than 1% of the total recovered fungal species from airborne eDNA studies [18], the majority of fungal spore viability cannot be assessed in this way.

An alternative method to determine eDNA viability is to use high-throughput sequencing and leverage the fact that active cells typically have higher concentrations of ribosomal RNA. Therefore, species with a high rRNA-to-DNA ratio are more likely to be active than those with a low abundance of rRNA. A study applying this principle to airborne eDNA identified a significant difference in the taxonomic abundance of DNA and RNA communities from the same sample, suggesting that not all the identified species were active [107]. However, this approach may be less applicable to organisms that disperse in a dormant state, such as spores, which remain viable but exhibit little or no rRNA activity.

Researchers are also exploring the possibility of interrogating raw nanopore data to determine if microbes are alive or dead based on the signal produced during sequencing [135]. Since nanopore technology sequences native DNA, it can detect DNA damage and modifications as DNA passes through the pore, represented as signals. Machine learning models interpret these signals, with a recent study training models on DNA from dead and live cells to identify how the signals differ in viable cells [135]. While still in the early stages, this approach could offer another method to assess the proportion of viable organisms in metagenomic samples.

Low biomass in the air

There is very little biomass present in the air and, therefore, low concentrations of DNA and, especially, RNA, to collect and sequence. A study of the SARS-CoV-2 virus found the concentration of its genome was just 0.87 copies per litre of air, which was judged insufficient for NGS methods but still suitable for Sanger sequencing. Another study collected bacteria and virus-like particles by size and found they were present at $\sim 10^5$ and 10^6 particles on average per cubic metre of outdoor air [136]. Airborne DNA concentrations also vary seasonally. A study, in Siberia, found a 170-fold increase in DNA yield in the summer compared to winter [43], though such large variations may not be typical in less extreme climates. Another study identified many more sequenced reads in spring compared to autumn, likely due to plant species growing, flowering and releasing pollen in the spring [137].

The concentration of DNA also varies by location: in areas such as Antarctica, there is very little DNA present in the air due to the presence of few living organisms. To overcome this limitation in one Antarctic study, samplers were operated for 2 weeks to collect 16,000 m³ of air. The DNA yield, measured by 16S concentration, was still found to be the same as negative controls [35].

Due to the low concentration of airborne DNA, some studies have refined pipelines to increase the concentration of DNA collected, allowing for shorter collection times. For example, one study demonstrated that improvements to the sampling, storage, extraction

and nucleic acid analysis steps increased DNA accumulation by 8–170-fold over other airborne eDNA studies [34]. Furthermore, a new DNA-isolation method from the MetaSUB project increased the DNA yield and reported diversity of airborne eDNA from subway samples, but not simple mocks [62]. The MetaSUB method uses a chemical lysis buffer and centrifugation process to remove DNA from the sampler's filter, followed by enzymatic and mechanical lysis to increase DNA yield.

Preventing contamination

Airborne eDNA sequencing is inherently at risk from contamination, as the low biomass means that any taxa introduced, even at low quantities, will be retained through standard analysis. Additionally, contamination could be added during the amplification processes necessary for metabarcoding [6]. Contamination is a known concern in eDNA studies, and as such, there are reviews on how to minimize both cross-contamination (between samples) and external contamination (from the laboratory, reagents or researchers) [138, 139].

External contamination can be minimized with sterile procedures and managed through the collection of negative controls from all steps of sampling and processing [140]. Negative controls are important as DNA extraction kits and other consumables may contain DNA, which will be present even under sterile conditions [59, 141]. Once sequenced, any taxa identified in negative controls can be further scrutinized in the analysis to determine the origin of the samples [81].

Human DNA can also contaminate samples, either from the researchers or others present in the sampling area. This type of contamination has been referred to as 'human genetic bycatch' and raises ethical concerns. For instance, one study was able to detect some haplotypes of people present in a room from 5 h of sampling [142]. In busy public areas, it will not be feasible to obtain consent from everyone present, so it is recommended that reads aligned to the human genome are removed from datasets before publication.

How does sequence data relate to species abundance?

A drawback of current eDNA analysis approaches is that the absolute concentration of taxa in the sample cannot be accurately inferred. In air studies, once DNA is extracted and sequenced, it is possible to calculate the concentration of DNA in the air. However, different taxa contain varying quantities of DNA, so there is no linear relationship between DNA concentration and taxon abundance as determined from sequence data.

A concern with eDNA datasets is their inherent compositional nature, meaning they reflect relative, rather than absolute, abundance. For example, a spike in one phylum's abundance can artificially reduce the apparent abundance of others, complicating the assessment of true abundance changes. This is especially problematic in airborne longitudinal studies, where shifts in relative proportions may not reflect actual increases or decreases in the abundance of taxa, but rather changes in the overall community composition (e.g. due to the release of pollen in spring/summer). For more details on compositionality in eDNA data, see [143].

One way to overcome this compositionality could be with spike-ins of synthetic DNAs at known concentrations. These have been used to provide semi-quantitative estimates of fungal DNA concentration in air [10].

Additionally, some evidence suggests a correlation between airborne eDNA data and species abundance. For example, a study monitoring agricultural pathogens with WGS quantified pathogen abundance changes over the growing season and correlated these with environmental factors [4]. Wind tunnel experiments have also shown that increasing the number of spores released from particular taxa leads to a higher relative abundance of that taxa in sequenced samples [4]. Therefore, relative abundance can serve as a proxy for airborne spore concentration.

With enough genomic data, it should be possible to estimate population sizes of detected taxa. In aquatic environments, genomic data has been successfully used for this purpose [144]. However, the DNA concentrations in airborne eDNA samples are currently too low to achieve the genomic coverage required for such analysis. As sampler sensitivity and sequencing technologies continue to improve, it may become feasible to collect and sequence enough airborne eDNA to estimate effective population sizes of taxa. This could open new possibilities for population-level studies of airborne taxa.

Microbial source identification

Biological particles can travel great distances depending on their size, wind speed and direction and other meteorological factors. For organisms that rely on air dispersion, such as spores and pollen, dispersal patterns have evolved over millennia. Some species remain near their origin due to favourable environmental conditions, while others travel far, possibly to reduce host density for pathogens. The factors controlling particle dispersion interact in different ways to affect particles along their trajectory before potential deposition onto an air sampler, making it exponentially difficult to ascertain the source of material using meteorological data.

Identifying the origin of airborne biological material is crucial for disease control and bioterrorism detection. For instance, if an air sampler detects foot-and-mouth disease DNA, pinpointing the farm of origin is critical for containment, a principle also

applicable to plant or human diseases. Many studies infer source locations of airborne eDNA based on detected taxa, such as wheat pathogens originating from nearby wheat fields. One study used light microscopy of Hirst traps in 38 locations to track *Alternaria* spores travelling from the Pannonian plain to Poland, identifying transport events through many factors including high *Alternaria* spore concentrations at the source and the detection of atypical spores at the destination [145]. This study also applied atmospheric footprint modelling to estimate air mass trajectories before deposition onto the samplers.

Atmospheric footprint modelling relies on particle dispersion models, such as the HYSPLIT (Hybrid-Single Particle Lagrangian Integrated Trajectory) model. HYSPLIT uses meteorological data to determine air mass movement and to infer the transport pathways of airborne particles [146]. However, real-world application is challenging due to variables like particle size, topography, wind speed and direction, humidity and precipitation. Despite these complexities, particle modelling has been used in airborne pollen and eDNA studies to elucidate the source of airborne taxa [131, 145, 147, 148]. Particle dispersion models can also be used to guide sampling location, the importance of which was discussed in Section 4.1.4.

Advances in aerobiology, including expanded sampling networks and refined dispersion modelling, are improving our ability to track microbial sources with greater accuracy. While precise pinpointing may remain difficult, integrating meteorological models with airborne eDNA monitoring will enhance source attribution and sampling strategies.

LOOKING TO THE FUTURE

The field of airborne eDNA analysis is rapidly advancing, moving beyond slower culture-based and microscopy methods towards real-time detection techniques. Within the next few years, ongoing technological advancements will likely open new avenues for airborne eDNA studies, enhancing our understanding of microbial communities and their environmental and health implications. This section highlights key future directions for air sampling, focusing on continuous monitoring, *in situ* testing, plant pathogen monitoring networks and improvements to low-input sequencing.

Proof-of-concept studies have shown that with improved technology, air samplers could enable real-time pathogen surveillance in agricultural fields, barns or high foot-traffic zones like airports, alerting farmers or authorities to potential threats [2, 23, 81]. Portable devices may also be used for personalized monitoring [37, 82]. Endangered species could also be monitored using non-invasive techniques [6, 20].

In healthcare settings, air samplers could help to detect and mitigate airborne disease spread in high-risk areas. For instance, deploying air monitors in wards for immunocompromised patients could enable rapid action when harmful bacteria are detected. Memon *et al.* provide an in-depth review summarizing the importance of indoor air sampling for public health and the available devices [149].

In situ testing in fields, hospitals and public transport is becoming more feasible with smaller hardware, lower power consumption and reduced computing requirements. A key challenge for *in situ* deployment is the automation of DNA processing and sequencing. While portable PCR and loop-mediated isothermal amplification technologies already exist, DNA extraction and library preparation are still largely confined to laboratories. Emerging devices like Oxford Nanopore Technologies' planned TraxION and Integra Biosciences' MIRO CANVAS represent a new generation of digital microfluidics capable of automating protocols, potentially enabling field deployment.

In situ sample testing will lead to faster diagnostics and rapid decision-making, eliminating delays associated with transport and laboratory processing. This is critical for responding to human health threats or plant pathogen outbreaks. Additionally, it removes the need for sample storage and transport, reducing the risk of deterioration and contamination.

Over the next decade, sequencing-based pathogen analysis could revolutionize agriculture. Fungal spore monitoring networks, already used in French vineyards [150], could be expanded to model and predict disease severity. These networks could also leverage existing air monitoring infrastructure [36], providing early warnings when emerging pathogens spread to new regions, an increasing concern due to a warming climate and increased globalization [151]. Air samplers could be deployed in warehouses storing fresh produce to monitor postharvest diseases, minimizing yield loss and preventing toxic chemical buildup in stored grain. In the future, *in situ* testing could even be integrated with precision agriculture, where air sampling data guides real-time application via direct injection mixture spray systems on tractors.

As technologies advance, with improved samplers, enhanced low-input sequencing protocols and reduced error rates, airborne eDNA analysis will become more precise, sensitive and scalable. This could enable the detection of antimicrobial, fungicide and herbicide resistant genes within pathogens [11, 81, 83], strain-level taxonomic identification [24, 152] and even SNP-based monitoring to measure effective population sizes, which could be especially valuable for tracking endangered species [144].

Interdisciplinary collaboration and technological innovation will be key to unlocking the full potential of airborne eDNA for real-time monitoring. Advances in sequencing accuracy, reference databases and source-tracking models will enhance reliability.

Additionally, addressing challenges to *in situ* processing, viability assessment and detection sensitivity will further expand its applications in health, agriculture and environmental monitoring.

Funding information

The authors acknowledge the support of the Biotechnology and Biological Sciences Research Council (BBSRC), part of UK Research and Innovation; Cross Institute Strategic Programme Grant Delivering Sustainable Wheat BBX011003/1 and its constituent work packages BBS/E/ER/230003B (DSW WP2 Delivering Resilience to Biotic Stress), BBS/E/ER/230003C (DSW WP4 Sustainable Data Frameworks for Wheat); and Earlham Institute Strategic Programme Grant Decoding Biodiversity BBX011089/1 and its constituent work package BBS/E/ER/230002A (Decode WP1 Development of Novel Experimental and Bioinformatic Tools for Genomic Diversity and Analysis). M.F.G.B. is supported by the UKRI BBSRC Norwich Research Park Biosciences Doctoral Training Partnership [Grant number BB/T008717/1].

Acknowledgements

The authors acknowledge the help of Pete Moore Fuller at Infohackit for his advice and feedback on the figures created.

Author contributions

M.F.G.B. led the drafting and editing of the manuscript. All other authors contributed to the drafting and editing. All authors approved the final manuscript.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

- Lednický JA, Shankar SN, Elbadry MA, Gibson JC, Alam MdM, et al. Collection of SARS-CoV-2 virus from the air of a clinic within a university student health care center and analyses of the viral genomic sequence. *Aerosol Air Qual Res* 2020;20:1167–1171.
- Haverkamp THA, Spilsberg B, Johannessen GS, Torp M, Sekse C. Detection and characterization of *Campylobacter* in air samples from poultry houses using shot-gun metagenomics – a pilot study. *BMC Microbiol* 2024;24:399.
- Lemos MSC de, Higa Junior MG, Paniago AMM, Melhem M de SC, Takahashi JPF, et al. *Aspergillus* in the indoor air of critical areas of a tertiary hospital in Brazil. *J Fungi* 2024;10:538.
- Giolai M, Verweij W, Martin S, Pearson N, Nicholson P, et al. Measuring air metagenomic diversity in an agricultural ecosystem. *Curr Biol* 2024;34:3778–3791.
- Huang S, Hu W, Chen J, Wu Z, Zhang D, et al. Overview of biological ice nucleating particles in the atmosphere. *Environ Int* 2021;146:106197.
- Lynggaard C, Bertelsen MF, Jensen CV, Johnson MS, Frøslev TG, et al. Airborne environmental DNA for terrestrial vertebrate community monitoring. *Curr Biol* 2022;32:701–707.
- Thuillet A-C, Morisot D, Renno J-F, Scarcelli N, Serret J, et al. Picturing plant biodiversity from airborne environmental DNA. *Genetics* 2024.
- Chappuis C, Tummon F, Clot B, Konzelmann T, Calpini B, et al. Automatic pollen monitoring: first insights from hourly data. *Aerobiologia* 2020;36:159–170.
- Banchi E, Ametrano CG, Tordoni E, Stanković D, Ongaro S, et al. Environmental DNA assessment of airborne plant and fungal seasonal diversity. *Sci Total Environ* 2020;738:140249.
- Ovaskainen O, Abrego N, Furneaux B, Hardwick B, Somervuo P, et al. Global spore sampling project: a global, standardized dataset of airborne fungal DNA. *Sci Data* 2024;11:561.
- Li L, Wang Q, Bi W, Hou J, Xue Y, et al. Municipal solid waste treatment system increases ambient airborne bacteria and antibiotic resistance genes. *Environ Sci Technol* 2020;54:3900–3908.
- Maki T, Hosaka K, Lee KC, Kawabata Y, Kajino M, et al. Vertical distribution of airborne microorganisms over forest environments: a potential source of ice-nucleating bioaerosols. *Atmos Environ* 2023;302:119726.
- Li K, Dong S, Wu Y, Yao M. Comparison of the biological content of air samples collected at ground level and at higher elevation. *Aerobiologia* 2010;26:233–244.
- Pollegioni P, Cardoni S, Mattioni C, Piredda R, Ristorini M, et al. Variability of airborne microbiome at different urban sites across seasons: a case study in Rome. *Front Environ Sci* 2023;11.
- Shi Y, Lai S, Liu Y, Gromov S, Zhang Y. Fungal aerosol diversity over the northern South China Sea: the influence of land and ocean. *JGR Atmospheres* 2022;127:e2021JD035213.
- Abrego N, Norros V, Halme P, Somervuo P, Ali-Kovero H, et al. Give me a sample of air and I will tell which species are found from your region: molecular identification of fungi from airborne spore samples. *Mol Ecol Resour* 2018;18:511–524.
- An C, Woo C, Yamamoto N. Introducing DNA-based methods to compare fungal microbiota and concentrations in indoor, outdoor, and personal air. *Aerobiologia* 2018;34:1–12.
- Mbareche H, Veillette M, Teertstra W, Kegel W, Bilodeau GJ, et al. Recovery of fungal cells from air samples: a tale of loss and gain. *Appl Environ Microbiol* 2019;85:e02941–18.
- Razzini K, Castrica M, Menchetti L, Maggi L, Negroni L, et al. SARS-CoV-2 RNA detection in the air and on surfaces in the COVID-19 ward of a hospital in Milan, Italy. *Sci Total Environ* 2020;742:140540.
- Clare EL, Economou CK, Faulkes CG, Gilbert JD, Bennett F, et al. eDNAir: proof of concept that animal DNA can be collected from air sampling. *PeerJ* 2021;9:e11030.
- Aguayo J, Husson C, Chancerel E, Fabreguettes O, Chandelier A, et al. Combining permanent aerobiological networks and molecular analyses for large-scale surveillance of forest fungal pathogens: a proof-of-concept. *Plant Pathol* 2021;70:181–194.
- Buters J, Clot B, Galán C, Gehrig R, Gilge S, et al. Automatic detection of airborne pollen: an overview. *Aerobiologia* 2024;40:13–37.
- West JS, Kimber RBE. Innovations in air sampling to detect plant pathogens. *Ann Appl Biol* 2015;166:4–17.
- Ferguson RMW, Garcia-Alcega S, Coulon F, Dumbrell AJ, Whitby C, et al. Bioaerosol biomonitoring: sampling optimization for molecular microbial ecology. *Mol Ecol Resour* 2019;19:672–690.
- Manibusan S, Mainelis G. Passive bioaerosol samplers: a complementary tool for bioaerosol research. A review. *J Aerosol Sci* 2022;163:105992.
- Check JC, Harkness RJ, Heger L, Sakalidis ML, Chilvers MI, et al. It's a trap! Part I: exploring the applications of rotating-arm impaction samplers in plant pathology. *Plant Dis* 2024;108:1910–1922.
- Mescioglou E, Paytan A, Mitchell BW, Griffin DW. Efficiency of bioaerosol samplers: a comparison study. *Aerobiologia* 2021;37:447–459.
- Jeong S -Y., Kim TG. Comparison of five membrane filters to collect bioaerosols for airborne microbiome analysis. *J Appl Microbiol* 2021;131:780–790.
- Hoisington AJ, Maestre JP, King MD, Siegel JA, Kinney KA. Impact of sampler selection on the characterization of the indoor

- microbiome via high-throughput sequencing. *Build Environ* 2014;80:274–282.
30. Pan M, Bonny TS, Loeb J, Jiang X, Lednicky JA, et al. Collection of viable aerosolized influenza virus and other respiratory viruses in a student health care center through water-based condensation growth. *mSphere* 2017;2:e00251-17.
 31. Raynor PC, Adesina A, Aboubakr HA, Yang M, Torremorell M, et al. Comparison of samplers collecting airborne influenza viruses: 1. Primarily impingers and cyclones. *PLoS One* 2021;16:e0244977.
 32. Abeykoon AMH, Poon M, Firestone SM, Stevenson MA, Wiethoelter AK, et al. Performance evaluation and validation of air samplers to detect aerosolized *Coxiella burnetii*. *Microbiol Spectr* 2022;10:e0065522.
 33. Polling M, Buij R, Laros I, de Groot GA. Continuous daily sampling of airborne eDNA detects all vertebrate species identified by camera traps. *Environ DNA* 2024;6:e591.
 34. Luhung I, Uchida A, Lim SBY, Gaultier NE, Kee C, et al. Experimental parameters defining ultra-low biomass bioaerosol analysis. *Npj Biofilms Microbiomes* 2021;7:37.
 35. Dommergue A, Amato P, Tignat-Perrier R, Magand O, Thollot A, et al. Methods to investigate the global atmospheric microbiome. *Front Microbiol* 2019;10.
 36. Littlefair JE, Allerton JJ, Brown AS, Butterfield DM, Robins C, et al. Air-quality networks collect environmental DNA with the potential to measure biodiversity at continental scales. *Curr Biol* 2023;33:R426–R428.
 37. Agranovski IE, Usachev EV. In-situ rapid bioaerosol detection in the ambient air by miniature multiplex PCR utilizing technique. *Atmos Environ* 2021;246:118147.
 38. Mahaffee WF, Margairaz F, Ulmer L, Bailey BN, Stoll R. Catching spores: linking epidemiology, pathogen biology, and physics to ground-based airborne inoculum monitoring. *Plant Dis* 2023;107:13–33.
 39. King P, Pham LK, Waltz S, Sphar D, Yamamoto RT, et al. Longitudinal metagenomic analysis of hospital air identifies clinically relevant microbes. *PLoS One* 2016;11:e0160124.
 40. Richardson M, Gottel N, Gilbert JA, Gordon J, Gandhi P, et al. Concurrent measurement of microbiome and allergens in the air of bedrooms of allergy disease patients in the Chicago area. *Microbiome* 2019;7:82.
 41. Birgand G, Peiffer-Smadja N, Fournier S, Kerneis S, Lescure F-X, et al. Assessment of air contamination by SARS-CoV-2 in hospital settings. *JAMA Netw Open* 2020;3:e2033232.
 42. Nicolaisen M, West JS, Sapkota R, Canning GGM, Schoen C, et al. Fungal communities including plant pathogens in near surface air are similar across Northwestern Europe. *Front Microbiol* 2017;8:1729.
 43. Gusareva ES, Gaultier NPE, Premkrishnan BNV, Kee C, Lim SBY, et al. Taxonomic composition and seasonal dynamics of the air microbiome in West Siberia. *Sci Rep* 2020;10:21515.
 44. Sullivan AR, Karlsson E, Svensson D, Brindefalk B, Villegas JA, et al. Airborne eDNA captures three decades of ecosystem biodiversity. *Ecology* 2023.
 45. Glawe DA. The powdery mildews: a review of the world's most familiar (yet poorly known) plant pathogens. *Annu Rev Phytopathol* 2008;46:27–51.
 46. Lagomarsino Oneto D, Golan J, Mazzino A, Pringle A, Seminara A. Timing of fungal spore release dictates survival during atmospheric transport. *Proc Natl Acad Sci USA* 2020;117:5134–5143.
 47. Jang GI, Hwang CY, Cho BC. Effects of heavy rainfall on the composition of airborne bacterial communities. *Front Environ Sci Eng* 2018;12:12.
 48. Niu M, Hu W, Cheng B, Wu L, Ren L, et al. Influence of rainfall on fungal aerobiota in the urban atmosphere over Tianjin, China: a case study. *Atmospheric Environ X* 2021;12:100137.
 49. Uetake J, Tobo Y, Uji Y, Hill TCJ, DeMott PJ, et al. Seasonal changes of airborne bacterial communities over Tokyo and influence of local meteorology. *Front Microbiol* 2019;10:1572.
 50. Franco Ortega S, Ferrocino I, Adams I, Silvestri S, Spadaro D, et al. Monitoring and surveillance of aerial mycobiota of rice paddy through DNA metabarcoding and qPCR. *J Fungi* 2020;6:372.
 51. Arigela R, Jose C, Gopalakrishnan S, Gunthe SS, Raghunathan R. Effect of relative humidity on passive spore release from substrate surfaces. *J Aerosol Sci* 2025;183:106477.
 52. Hu Z, Liu H, Zhang H, Zhang X, Zhou M, et al. Temporal discrepancy of airborne total bacteria and pathogenic bacteria between day and night. *Environ Res* 2020;186:109540.
 53. Clauss M. Particle size distribution of airborne micro-organisms in the environment—a review. *Landbauforsch Volkenrode* 2015;65:77–100.
 54. Tsuda A, Henry FS, Butler JP. Particle transport and deposition: basic physics of particle kinetics. *Compr Physiol* 2013;3:1437–1471.
 55. Alsved M, Fraenkel C-J, Bohgard M, Widell A, Söderlund-Strand A, et al. Sources of airborne norovirus in hospital outbreaks. *Clin Infect Dis* 2020;70:2023–2028.
 56. Yamamoto N, Bibby K, Qian J, Hospodsky D, Rismani-Yazdi H, et al. Particle-size distributions and seasonal diversity of allergenic and pathogenic fungi in outdoor air. *ISME J* 2012;6:1801–1811.
 57. Ferguson RMW, Neath CEE, Nasir ZA, Garcia-Alcega S, Tyrrel S, et al. Size fractionation of bioaerosol emissions from green-waste composting. *Environ Int* 2021;147:106327.
 58. Jaing C, Thissen J, Morrison M, Dillon MB, Waters SM, et al. Sierra Nevada sweep: metagenomic measurements of bioaerosols vertically distributed across the troposphere. *Sci Rep* 2020;10:12399.
 59. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 2014;12:87.
 60. Brooks JP, Edwards DJ, Harwich MD Jr, Rivera MC, Fettweis JM, et al. The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiol* 2015;15:66.
 61. Banchi E, Pallavicini A, Muggia L. Relevance of plant and fungal DNA metabarcoding in aerobiology. *Aerobiologia* 2020;36:9–23.
 62. Boïfot KO, Gohli J, Moen LV, Dybwad M. Performance evaluation of a new custom, multi-component DNA isolation method optimized for use in shotgun metagenomic sequencing-based aerosol microbiome research. *Environ Microbiome* 2020;15:1.
 63. Quick J. The “Three Peaks” faecal DNA extraction method for long-read sequencing v2. 2019. 10.17504/protocols.io.7rsh6e
 64. Brodie EL, DeSantis TZ, Parker JPM, Zubieta IX, Piceno YM, et al. Urban aerosols harbor diverse and dynamic bacterial populations. *Proc Natl Acad Sci USA* 2007;104:299–304.
 65. Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, et al. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol* 2014;12:e1001889.
 66. Chen W, Hambleton S, Seifert KA, Carisse O, Diarra MS, et al. Assessing performance of spore samplers in monitoring aeromycobiota and fungal plant pathogen diversity in Canada. *Appl Environ Microbiol* 2018;84:e02601-17.
 67. Scibetta S, Schena L, Abdelfattah A, Pangallo S, Cacciola SO. Selection and experimental evaluation of universal primers to study the fungal microbiome of higher plants. *Phytobiomes J* 2018;2:225–236.
 68. Bakker MG. A fungal mock community control for amplicon sequencing experiments. *Mol Ecol Resour* 2018;18:541–556.
 69. Low L, Fuentes-Utrilla P, Hodson J, O'Neil JD, Rossiter AE, et al. Evaluation of full-length nanopore 16S sequencing for detection of pathogens in microbial keratitis. *PeerJ* 2021;9:e10778.

70. Schloss PD, Jenior ML, Koumpouras CC, Westcott SL, Highlander SK. Sequencing 16S rRNA gene fragments using the PacBio SMRT DNA sequencing system. *PeerJ* 2016;4:e1869.
71. Cuber P, Chooneea D, Geeves C, Salatino S, Creedy TJ, et al. Comparing the accuracy and efficiency of third generation sequencing technologies, Oxford Nanopore Technologies, and Pacific Biosciences, for DNA barcode sequencing applications. *Ecol Genet Genomics* 2023;28:100181.
72. Santos A, van Aerle R, Barrientos L, Martinez-Urtaza J. Computational methods for 16S metabarcoding studies using Nanopore sequencing data. *Comput Struct Biotechnol J* 2020;18:296–305.
73. Silverman JD, Bloom RJ, Jiang S, Durand HK, Dallow E, et al. Measuring and mitigating PCR bias in microbiota datasets. *PLoS Comput Biol* 2021;17:e1009113.
74. Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ Microbiol* 2016;18:1403–1414.
75. Kelly RP, Shelton AO, Gallego R. Understanding PCR processes to draw meaningful conclusions from environmental dna studies. *Sci Rep* 2019;9:12133.
76. Forry SP, Servetas SL, Kralj JG, Soh K, Hadjithomas M, et al. Variability and bias in microbiome metagenomic sequencing: an interlaboratory study comparing experimental protocols. *Sci Rep* 2024;14:9785.
77. Peng W, Li X, Wang C, Cao H, Cui Z. Metagenome complexity and template length are the main causes of bias in PCR-based bacteria community analysis. *J Basic Microbiol* 2018;58:987–997.
78. Větrovský T, Baldrian P. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS One* 2013;8:e57923.
79. Schloss PD. Identifying and overcoming threats to reproducibility, replicability, robustness, and generalizability in microbiome research. *mBio* 2018;9:00525–18.
80. Regueira-Iglesias A, Balsa-Castro C, Blanco-Pintos T, Tomás I. Critical review of 16S rRNA gene sequencing workflow in microbiome studies: from primer selection to advanced data analysis. *Mol Oral Microbiol* 2023;38:347–399.
81. Leung MHY, Tong X, Bøifot KO, Bezdán D, Butler DJ, et al. Characterization of the public transit air microbiome and resistome reveals geographical specificity. *Microbiome* 2021;9:112.
82. Jiang C, Zhang X, Gao P, Chen Q, Snyder M. Decoding personal biotic and abiotic airborne exposome. *Nat Protoc* 2021;16:1129–1151.
83. Qin N, Liang P, Wu C, Wang G, Xu Q, et al. Longitudinal survey of microbiome associated with particulate matter in a megacity. *Genome Biol* 2020;21:55.
84. Heavens D, Chooneea D, Giolai M, Cuber P, Aanstad P, et al. How low can you go? Driving down the DNA input requirements for nanopore sequencing. *Genomics* 2021.
85. Ruppert KM, Kline RJ, Rahman MS. Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: a systematic review in methods, monitoring, and applications of global eDNA. *Glob Ecol Conserv* 2019;17:e00547.
86. Yang L, Chen J. Benchmarking differential abundance analysis methods for correlated microbiome sequencing data. *Brief Bioinform* 2023;24:bbac607.
87. Boshuizen HC, Te Beest DE. Pitfalls in the statistical analysis of microbiome amplicon sequencing data. *Mol Ecol Resour* 2023;23:539–548.
88. Kim C, Pongpanich M, Porntaveetus T. Unraveling metagenomics through long-read sequencing: a comprehensive review. *J Transl Med* 2024;22:111.
89. Liu Y-X, Qin Y, Chen T, Lu M, Qian X, et al. A practical guide to amplicon and metagenomic analysis of microbiome data. *Protein Cell* 2021;12:315–330.
90. Hemstrom W, Grummer JA, Luikart G, Christie MR. Next-generation data filtering in the genomics era. *Nat Rev Genet* 2024;25:750–767.
91. Ayling M, Clark MD, Leggett RM. New approaches for metagenome assembly with short reads. *Brief Bioinform* 2020;21:584–594.
92. Pilo P, Lawless C, Tiley AMM, Karki SJ, Burke JI, et al. Comparison of microscopic and metagenomic approaches to identify cereal pathogens and track fungal spore release in the field. *Front Plant Sci* 2022.
93. Ghurye JS, Cepeda-Espinoza V, Pop M. Metagenomic assembly: overview, challenges and applications. *Yale J Biol Med* 2016;89:353–362.
94. Mallawaarachchi V, Wickramarachchi A, Xue H, Papudeshi B, Grigson SR, et al. Solving genomic puzzles: computational methods for metagenomic binning. *Brief Bioinform* 2024;25:bbae372.
95. Meyer F, Fritz A, Deng Z-L, Koslicki D, Lesker TR, et al. Critical assessment of metagenome interpretation: the second round of challenges. *Nat Methods* 2022;19:429–440.
96. Hakimzadeh A, Abdala Asbun A, Albanese D, Bernard M, Buchner D, et al. A pile of pipelines: an overview of the bioinformatics software for metabarcoding data analyses. *Mol Ecol Resour* 2024;24:e13847.
97. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410.
98. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol* 2019;20:257.
99. Blanco-Míguez A, Beghini F, Cumbo F, McIver LJ, Thompson KN, et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species with MetaPhlan 4. *Bioinformatics* 2022.
100. Smith RH, Glendinning L, Walker AW, Watson M. Investigating the impact of database choice on the accuracy of metagenomic read classification for the rumen microbiome. *Anim Microbiome* 2022;4:57.
101. Nasko DJ, Koren S, Phillippy AM, Treangen TJ. RefSeq database growth influences the accuracy of k-mer-based lowest common ancestor species identification. *Genome Biol* 2018;19:165.
102. Chorlton SD. Ten common issues with reference sequence databases and how to mitigate them. *Front Bioinform* 2024;4.
103. Reska T, Pozdniakova S, Borràs S, Perlas A, Sauerborn E, et al. Air monitoring by nanopore sequencing. *ISME Commun* 2024;4:ycac099.
104. Nilsson RH, Larsson K-H, Taylor AFS, Bengtsson-Palme J, Jeppesen TS, et al. The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Res* 2019;47:D259–D264.
105. Apangu GP, Frisk CA, Petch GM, Muggia L, Pallavicini A, et al. Environmental DNA reveals diversity and abundance of Alternaria species in neighbouring heterogeneous landscapes in Worcester, UK. *Aerobiologia* 2022;38:457–481.
106. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* 2007;35:7188–7196.
107. Gomez-Silvan C, Leung MHY, Grue KA, Kaur R, Tong X, et al. A comparison of methods used to unveil the genetic and metabolic pool in the built environment. *Microbiome* 2018;6:71.
108. Mayol E, Arrieta JM, Jiménez MA, Martínez-Asensio A, Garcías-Bonet N, et al. Long-range transport of airborne microbes over the global tropical and subtropical ocean. *Nat Commun* 2017;8:201.
109. Prussin AJ II, Torres PJ, Shimashita J, Head SR, Bibby KJ, et al. Seasonal dynamics of DNA and RNA viral bioaerosol communities in a daycare center. *Microbiome* 2019;7:53.
110. Urban M, Cuzick A, Seager J, Wood V, Rutherford K, et al. PHI-base: the pathogen–host interactions database. *Nucleic Acids Res* 2020;48:D613–20.

111. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res* 2017;45:D566–D573.
112. Yin X, Jiang X-T, Chai B, Li L, Yang Y, et al. ARGs-OAP v2.0 with an expanded SARG database and hidden Markov models for enhancement characterization and quantification of antibiotic resistance genes in environmental metagenomes. *Bioinformatics* 2018;34:2263–2270.
113. He P, Wu Y, Huang W, Wu X, Lv J, et al. Characteristics of and variation in airborne ARGs among urban hospitals and adjacent urban and suburban communities: a metagenomic approach. *Environ Int* 2020;139:105625.
114. Song L, Ma J, Jiang G, Wang C, Zhang Y, et al. Comparison of airborne antibiotic resistance genes in the chicken farm during winter and summer. *Indoor Air* 2024;2024:1707863.
115. Nguyen NH, Song Z, Bates ST, Branco S, Tedersoo L, et al. FUNGuild: an open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecol* 2016;20:241–248.
116. Cao Y, Yu X, Ju F, Zhan H, Jiang B, et al. Airborne bacterial community diversity, source and function along the Antarctic Coast. *Sci Total Environ* 2021;765:142700.
117. Jiang X, Wang C, Guo J, Hou J, Guo X, et al. Global Meta-analysis of Airborne Bacterial Communities and Associations with Anthropogenic Activities. *Environ Sci Technol* 2022;56:9891–9902.
118. Yan X, Ma J, Chen X, Lei M, Li T, et al. Characteristics of airborne bacterial communities and antibiotic resistance genes under different air quality levels. *Environ Int* 2022;161:107127.
119. Feng G, Xie T, Wang X, Bai J, Tang L, et al. Metagenomic analysis of microbial community and function involved in cd-contaminated soil. *BMC Microbiol* 2018;18:11.
120. Bengtsson-Palme J, Alm Rosenblad M, Molin M, Blomberg A. Metagenomics reveals that detoxification systems are under-represented in marine bacterial communities. *BMC Genom* 2014;15:749.
121. Wen T, Niu G, Chen T, Shen Q, Yuan J, et al. The best practice for microbiome analysis using R. *Protein Cell* 2023;14:713–725.
122. Pan AY. Statistical analysis of microbiome data: the challenge of sparsity. *Curr Opin Endocr Metab Res* 2021;19:35–40.
123. Lutz KC, Jiang S, Neugent ML, De Nisco NJ, Zhan X, et al. A survey of statistical methods for microbiome data analysis. *Front Appl Math Stat* 2022;8:884810.
124. Wieczorek TM, Jørgensen LN, Hansen AL, Munk L, Justesen AF. Early detection of sugar beet pathogen *Ramularia beticola* in leaf and air samples using qPCR. *Eur J Plant Pathol* 2014;138:775–785.
125. Ginn O, Rocha-Melogno L, Bivins A, Lowry S, Cardelino M, et al. Detection and quantification of enteric pathogens in aerosols near open wastewater canals in cities with poor sanitation. *Environ Sci Technol* 2021;55:14758–14771.
126. Van der Heyden H, Dutilleul P, Charron J-B, Bilodeau GJ, Carisse O. Monitoring airborne inoculum for improved plant disease management. A review. *Agron Sustain Dev* 2021;41:40.
127. Bowers RM, Lauber CL, Wiedinmyer C, Hamady M, Hallar AG, et al. Characterization of airborne microbial communities at a high-elevation site and their potential to act as atmospheric ice nuclei. *Appl Environ Microbiol* 2009;75:5121–5130.
128. Folloni S, Kagkli D-M, Rajcevic B, Guimarães NCC, Van Droogenbroeck B, et al. Detection of airborne genetically modified maize pollen by real-time PCR. *Mol Ecol Resour* 2012;12:810–821.
129. Drautz-Moses DI, Luhung I, Gusareva ES, Kee C, Gaultier NE, et al. Vertical stratification of the air microbiome in the lower troposphere. *Proc Natl Acad Sci USA* 2022;119:e2117293119.
130. Song H, Crawford I, Lloyd J, Robinson C, Boothman C, et al. Airborne bacterial and eukaryotic community structure across the United Kingdom revealed by high-throughput sequencing. *Atmosphere* 2020;11:802.
131. Métris KL, Métris J. Aircraft surveys for air eDNA: probing biodiversity in the sky. *PeerJ* 2023;11:e15171.
132. Kobziar LN, Vuono D, Moore R, Christner BC, Dean T, et al. Wild-land fire smoke alters the composition, diversity, and potential atmospheric function of microbial life in the aerobiome. *ISME Commun* 2022;2:8.
133. Gould S, Atkinson B, Onianwa O, Spencer A, Furneaux J, et al. Air and surface sampling for monkeypox virus in a UK hospital: an observational study. *Lancet Microbe* 2022;3:e904–e911.
134. Ingram RJ, Ludwig HD, Scherm H. Epidemiology of *Exobasidium* leaf and fruit spot of rabbiteye blueberry: pathogen overwintering, primary infection, and disease progression on leaves and fruit. *Plant Dis* 2019;103:1293–1301.
135. Urel H, Benassou S, Reska T, Marti H, Rayo E, et al. Nanopore- and AI-empowered metagenomic viability inference. *Bioinformatics*. DOI: 10.1101/2024.06.10.598221
136. Prussin AJ, Garcia EB, Marr LC. Total virus and bacteria concentrations in indoor and outdoor air. *Environ Sci Technol Lett* 2015;2:84–88.
137. Johnson MD, Cox RD, Grisham BA, Lucia D, Barnes MA. Airborne eDNA reflects human activity and seasonal changes on a landscape scale. *Front Environ Sci* 2021;8:276.
138. Cando-Dumancela C, Liddicoat C, McLeod D, Young JM, Breed MF. A guide to minimize contamination issues in microbiome restoration studies. *Restor Ecol* 2021;29:e13358.
139. Sepulveda AJ, Hutchins PR, Forstchen M, Mckeefry MN, Swigris AM. The elephant in the lab (and field): contamination in aquatic environmental DNA studies. *Front Ecol Evol* 2020;8.
140. Jiang W, Liang P, Wang B, Fang J, Lang J, et al. Optimized DNA extraction and metagenomic sequencing of airborne microbial communities. *Nat Protoc* 2015;10:768–779.
141. Weyrich LS, Farrer AG, Eisenhofer R, Arriola LA, Young J, et al. Laboratory contamination over time during low-biomass sample analysis. *Mol Ecol Resour* 2019;19:982–996.
142. Whitmore L, McCauley M, Farrell JA, Stammnitz MR, Koda SA, et al. Inadvertent human genomic bycatch and intentional capture raise beneficial applications and ethical concerns with environmental DNA. *Nat Ecol Evol* 2023;7:873–888.
143. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: and this is not optional. *Front Microbiol* 2017;8:2224.
144. Sigsgaard EE, Jensen MR, Winkelmann IE, Møller PR, Hansen MM, et al. Population-level inferences from environmental DNA—current status and future perspectives. *Evol Appl* 2020;13:245–262.
145. Gwirling Ł, Magyar D, Chtopek K, Grinn-Gofroń A, Gwiązdowska J, et al. Bioaerosols on the atmospheric super highway: an example of long distance transport of *Alternaria* spores from the Pannonian Plain to Poland. *Sci Total Environ* 2022;819:153148.
146. Stein AF, Draxler RR, Rolph GD, Stunder BJB, Cohen MD, et al. NOAA's HYSPLIT atmospheric transport and dispersion modeling system. *Bull Am Meteorol Soc* 2015;96:2059–2077.
147. Hanson MC, Petch GM, Ottosen T-B, Skjøth CA. Climate change impact on fungi in the atmospheric microbiome. *Sci Total Environ* 2022;830:154491.
148. Chen W, Newlands N, Hambleton S, Laroche A, Davoodi SM, et al. Optimizing an integrated biovigilance toolbox to study the spatial distribution and dynamic changes of airborne mycobiota, with a focus on cereal rust fungi in western Canada. *Mol Ecol Resour* 2024;24:e13983.
149. Memon R, Niazi JH, Qureshi A. Biosensors for detection of airborne pathogenic fungal spores: a review. *Nanoscale* 2024;16:15419–15445.
150. Laurent B, Douillet A, Beslay A, Bordes J, Delmotte F, et al. The VISA network: a collaborative project between research institutes and vineyard owners to create the first epidemiological

- monitoring network of downy mildew epidemic based on aerial spore capture. *BIO Web Conf* 2022;50:04007.
151. Singh BK, Delgado-Baquerizo M, Egidi E, Guirado E, Leach JE, et al. Climate change impacts on plant pathogens, food security and paths forward. *Nat Rev Microbiol* 2023;21:640–656.
 152. Huffman JA, Perring AE, Savage NJ, Clot B, Crouzy B, et al. Real-time sensing of bioaerosols: review and current perspectives. *Aerosol Sci Technol* 2020;54:465–495.
 153. Taş N, de Jong AE, Li Y, Trubl G, Xue Y, et al. Metagenomic tools in microbial ecology research. *Curr Opin Biotechnol* 2021;67:184–191.
 154. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 2011;17:10.
 155. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018;34:i884–i890.
 156. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30:2114–2120.
 157. Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes* 2016;9:88.
 158. Leggett RM, Clavijo BJ, Clissold L, Clark MD, Caccamo M. NextClip: an analysis and read preparation tool for Nextera long mate pair libraries. *Bioinformatics* 2014;30:566–568.
 159. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 2012;28:3150–3152.
 160. Li H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* 2021;37:4572–4574.
 161. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009;10:R25.
 162. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate illumina paired-end read merger | bioinformatics | oxford academic. *Bioinformatics* 2014;30:614–620.
 163. Yang C, Chowdhury D, Zhang Z, Cheung WK, Lu A, et al. A review of computational tools for generating metagenome-assembled genomes from metagenomic sequencing data. *Comput Struct Biotechnol J* 2021;19:6301–6314.
 164. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 2015;31:1674–1676.
 165. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 2017;27:824–834.
 166. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, et al. metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat Methods* 2020;17:1103–1110.
 167. Benoit G, Raguideau S, James R, Phillippy AM, Chikhi R, et al. High-quality metagenome assembly from long accurate reads with metaMDBG. *Nat Biotechnol* 2024;42:1378–1383.
 168. Rho M, Tang H, Ye Y. FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res* 2010;38:e191.
 169. Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 2015;3:e1165.
 170. Wu Y-W, Tang Y-H, Tringe SG, Simmons BA, Singer SW. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* 2014;2:26.
 171. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, et al. Binning metagenomic contigs by coverage and composition. *Nat Methods* 2014;11:1144–1146.
 172. de Hoon MJL, Imoto S, Nolan J, Miyano S. Open source clustering software. *Bioinformatics* 2004;20:1453–1454.
 173. Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 2018;6:158.
 174. Terrón-Camero LC, Gordillo-González F, Salas-Espejo E, Andrés-León E. Comparison of metagenomics and metatranscriptomics tools: a guide to making the right choice. *Genes* 2022;13:2280.
 175. Li H, Durbin R. Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics* 2010;26:589–595.
 176. Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res* 2016;26:1721–1729.
 177. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol* 2016;17:132.
 178. Franzosa EA, McIver LJ, Rahnava G, Thompson LR, Schirmer M, et al. Species-level functional profiling of metagenomes and metatranscriptomes. *Nat Methods* 2018;15:962–968.
 179. Kaminski J, Gibson MK, Franzosa EA, Segata N, Dantas G, et al. High-specificity targeted functional profiling in microbial communities with ShortBRED. *PLoS Comput Biol* 2015;11:e1004557.
 180. Wemheuer F, Taylor JA, Daniel R, Johnston E, Meinicke P, et al. Tax4Fun2: prediction of habitat-specific functional profiles and functional redundancy based on 16S rRNA gene sequences. *Environ Microbiome* 2020;15:11.
 181. Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, et al. PICRUSt2 for prediction of metagenome functions. *Nat Biotechnol* 2020;38:685–688.
 182. Oksanen J, Simpson GL, Blanchet FG, Kindt R, Legendre P, et al. Vegan: community ecology package. 2024. 10.32614/CRAN.package.vegan
 183. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 2013;8:e61217.
 184. Valero-Mora PM. Ggplot2: elegant graphics for data analysis. *J Stat Softw* 2010;35:1–3.
 185. Herbig A, Maixner F, Bos KI, Zink A, Krause J, et al. MALT: fast alignment and analysis of metagenomic DNA sequence data applied to the Tyrolean Iceman. *Bioinformatics* 2016.
 186. Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. *Genome Res* 2007;17:377–386.
 187. Peel N, Martin S, Heavens D, Yu DW, Clark MD, et al. MARTi: a real-time analysis and visualisation tool for nanopore metagenomics. *bioRxiv*
 188. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, et al. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 2016;13:581–583.
 189. Sharma VK, Kumar N, Prakash T, Taylor TD. Fast and accurate taxonomic assignments of metagenomic sequences using MetaBin. *PLoS One* 2012;7:e34030.